170-WP-022-001

Final Report on CIP Prototype Study Engineering Support Directive 60

White Paper

White Paper--Not intended for formal review or government approval.

October 1999

Prepared Under Contract NAS5-60000

RESPONSIBLE AUTHOR

Ananth Rao /s/ 10/29/99

Ananth Rao, Systems Engineer
EOSDIS Core System Project

Date

SUBMITTED BY

Carroll A. Hood /s/ 11/1/99

Carroll Hood, Lead Engineer
EOSDIS Core System Project

Date

Raytheon Systems Company
Upper Marlboro, Maryland

This page intentionally left blank.

Abstract

This report describes the design and current implementation state of prototyping effort that was funded by Engineering Support Directive (ESD) 60. The period of performance covered by this ESD was from November 1, 1998 to October 31, 1999. The main focus of this effort was to allow interoperability between the Interoperable Catalog System (ICS) domain and other EO domains. The EO domains that were of interest, and were used for interoperability testing were ECS and IMS. Work covered by this ESD, was augmented by the lessons learned and prototyping efforts from an earlier Engineering Support Directive – ESD 38.

This report includes:

- Introduction, which includes the purpose and objectives of the prototyping effort;
- Results of implementation;
- Documentation generated during this effort;
- Concluding remarks and possible follow on work; and
- Lessons learned from this effort

Keywords: CIP, ECS, Gateway, IMS, INFEO, Interoperability, Prototype, Translator

This page intentionally left blank.

Contents

4. Lessons Learned and Future Work

List of Figures

List of Tables

Abbreviations and Acronyms

1. Introduction

1.1 Purpose

This report describes the prototyping efforts covered under ESD 60. The objectives of this effort were to provide interoperability between the ICS domain and the ECS and IMS domains. Two task threads were chosen for demonstration purposes:

- ICS Clients accessing ECS services and
- IMS Clients accessing ICS services

This document describes the design and implementation/configuration of the major components that constituted the above two tasks, as well as describing the lessons learned as a result of this effort. Possible follow-on tasks to this ESD are also described.

1.2 Acknowledgements

This document was developed with input from Ajay Gupte, and Janet Hylton, formerly of Raytheon Systems Company, and Eric S. Martin of Raytheon Systems Company.

1.3 References

References and applicable documents:

[DD-1] CIP-ECS Catalog Translator Design Document, Raytheon Systems Co. White Paper, October 1998.

[CP-1] Catalogue Interoperability Protocol (CIP) Specification – Release B [CEOS/WGISS/PTT/CIP-B, Issue 2.2] January 22, 1997.

[CP-2] Catalogue Interoperability Protocol (CIP) Specification – Release B [CEOS/WGISS/PTT/CIP-B, Issue 2.4] June 1998.

[SD-1] Interoperable Catalogue System (ICS) System Design Document – [CEOS/WGISS/PTT/SDD, Issue1.2] March 1997.

[IN-1] CEO INFEO Architecture Overview Document [CEO/ES/WP2300/F-ADD-111, Issue 1.0] January 1998.

[IN-2] CEO System Requirements and Architecture Document [CEO/ES/WP2200/SR-2/182, Revision 1] July 1996.

[EG-1] EOS Data Gateway [Version 2.3.2] Web Gateway Design Document, October 18, 1999.

[IG-1] Interoperable Catalog System Gateway Prototype Design Document [170-WP-015-001], August 1999.

[OH-1] Catalog Interoperability Protocol OHS Translator Design Document, Raytheon Systems Co. White Paper, August 1997.

1.4 Organization

This report is organized as follows: Section I states the main objectives of this ESD, section 2 describes the high level design of the components constituting the prototype, implementation and configuration issues are described in section 3 and section 4 states lessons learned from this effort as well as possible follow-on work to this ESD.

2. Prototype Design

2.1 ICS/ECS Gateway

The purpose of this subtask was to show that Clients in the ICS domain could access ECS meta-data holdings, both at a collection level as well as at a product level. The ICS framework and architecture are described in [SD-1]. The major design goals of this subtask were:

- Be a node in the ICS as defined by the ICS System Design Document [SD-1].
- Conform to the CIP Specification Version 2.2 [CP-1], with INFEO modifications [IN-2].
- Provide a migration path to upgrade the system to CIP Specification Version 2.4 [CP-2].
- Re-use existing prototype software
- Modular components
- Provide Flexibility for expansion

The original thinking for this task called for the design and implementation of a Catalog Translator which would interface with a Retrieval Manager on one side and with ECS servers on the other, translating service and protocol requests originating from ICS. A detailed design document with this approach appears in [DD-1]. This approach was however subsequently modified to re-use additional INFEO components (IMS Facility) and the V0 Gateway of ECS, and was deemed to be the quickest path to providing access to ECS.

2.1.1 System Architecture

The configuration used for providing access to ECS product level metadata is shown in Figure 2.1. Web Clients access the INFEO system (See Section 3.XX for a description of INFEO components) to search for ECS metadata holdings. CIP *Collection Search* requests are serviced by the INFEO Collections Database, populated with ECS Collection level metadata. CIP *Product Search* requests are routed to the INFEO IMS Facility, which maps CIP attributes into IMS attributes and converts the request to IMS format (ODL), which then forwards the request to the V0 Gateway. The V0 Gateway is the primary interface point into ECS services and converts this ODL request into an ECS request, mapping IMS attributes into ECS attributes along the way, and accesses the Science Data Server for product level metadata. Note that cross-DAAC searching facility is provided by the IMS Facility which allows search requests to be forwarded to multiple V0 servers and which then takes the responsibility of collating the results from different archive centers before returning them to the client.

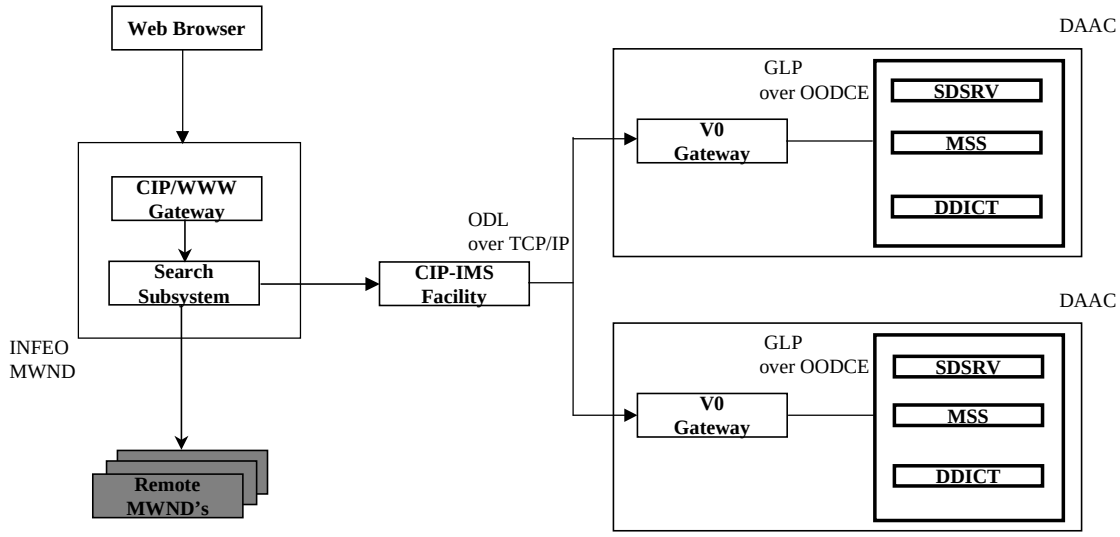

Figure 2-1. ICS-ECS Gateway Architecture

2.2 IMS/ICS Gateway

The purpose behind this subtask was to show interoperability between the IMS and ICS domains. The primary design goals of the gateway are as indicated in Section 2.1 above. In addition, the following objectives were required to be met:

- Conform to the latest version of the IMS Web gateway
- Provide Directory and Inventory Search capabilities

2.2.1 System Architecture

The architecture of the IMS/ICS Gateway system is shown in Figure 2-2. IMS clients accessing the IMS Web Gateway, wishing to search ICS holdings are directed to the IMS/ICS Gateway component which maps IMS attributes in the search request into CIP attributes and converts the incoming ODL request into CIP and forwards it to the INFEO system for execution. This scheme requires that the GCMD contain a DIF entry for each CIP Collection, and which is used to locate the collection of interest in the ICS domain.

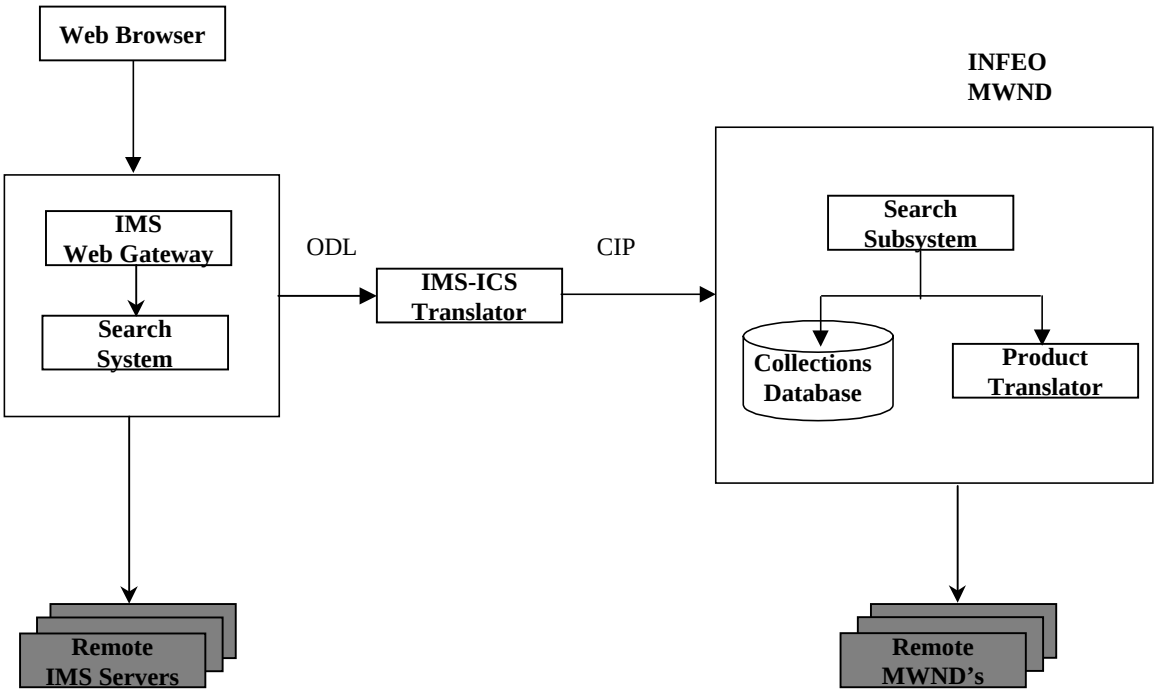

Figure 2-2. IMS-ICS Gateway Architecture

This page intentionally left blank.

3. Prototype Implementation/Configuration

3.1 ICS/ECS Gateway

The following components constitute the ICS/ECS Gateway.

- INFEO
- IMS Facility
- V0 Gateway

These subsystems are described in the following sections

3.1.1 INFEO

The INFEO system is a reference implementation of the ICS system, conforming to CIP Specification 2.2, with modifications (primarily in interpretation of the ICS Specification documents). A complete description of the system is described in references [IN-1] and [IN-2]. The software was licensed from the Joint Research Council of the European Space Agency, and was subsequently installed and configured on a local configuration machine.

The core components of the INFEO system are:

- Middleware Node (MWND), the user access point to data, information and services in INFEO. The MWND provides services for advertising, discovering and searching data and information at the MWND and at remote catalogs.
- Monitoring and Control Facility (MCF) for maintaining overall system management services.
- Provider gateways for facilitating access to data or bibliographic catalogs

The role of a MWND is to provide an access point for providers and customers. The following services are provided:

- Advertisement/Announcement services – supports querying and browsing of advertisements for EO data, documents and querying and browsing of announcements.
- Bibliographic Search service – querying of local inventory libraries and remote compatible library database
- Data Search service – querying of local CIP collections and remote CIP collections and product queries
- User Profile service – registration and user maintenance of registration details

- Data Dictionary service – management and provision of a system wide data dictionary
- Help Service – query, browse and display of help data about the use of the system
- Log and Maintenance service – internal system management reporting and analysis.

The role of the Monitoring and Control Facility is to co-ordinate the operation of the system. It controls the overall on-line help and data dictionary data and provides analysis and reporting tools for monitoring and coordination.

Provider Gateways allows connections to the INFEO system by small data providers and make their data holdings available to INFEO users in general. Currently supported provider gateways are:

- CIP/ODBC Gateway
- GILS/ODBC Gateway
- WAIS/ODBC Gateway
- CIP/IMS Facility

The aim of the provider gateways is to transform the incoming protocol from INFEO namely CIP (or other Z39.50 profiles such as GILS and WAIS), into a non-database dependent format, or to convert from CIP to the IMS protocol, thus allowing a wide range of providers to advertise their datasets through the system.

3.1.1.1 INFEO Configuration & Installation

This section describes the steps followed in configuring and installing the MWND system. Also described are changes that were made to work in the development environment. The primary subsystem of interest was the Search Subsystem, the MCF was not installed. The installation/configuration process entailed three major subtasks:

- Compilation/Configuration of the MWND software system
- Installation of the required COTS packages (Oracle Enterprise Server 7.3.3 & Developer 2000 1.3.2) and
- Population of the Collections Database with ECS specific collection level metadata.

The first two were straight forward albeit time consuming due to the fact that the INFEO delivered software system was a snapshot of the actual system and came with interim instructions on it's configuration.

Population of the Collections Database was much more time consuming, since it involved, first mapping CIP attributes into ECS (and vice-versa), obtaining the actual collection level metadata values (which are spread over a large number of ECS specific databases and subsystems) and finally, the writing of a series of filters to convert the data from one format to another in order to allow database population.

Issues encountered during INFEO configuration/installation

- Documentation Issues

  As mentioned earlier, the delivered system was a snapshot of the actual system and the system came with interim instructions on compilation and configuration.

- Configuration Issues

  The INFEO system depends on a large number of (Unix) environment variables to configure the system. These environment variables were set from a large number of shell configuration files, some of which were reset to different values from different subsystems. The problems were twofold: first a dependence on the Unix *ksh* shell interpreter and second, inconsistent setting of these environment variables from different subsystems, leading to serialization type errors depending on which subsystem's servers were started first.

- Bug Fixes

  The following changes were made to the INFEO code to fix perceived problems at run-time. Note that these problems might have been fixed in subsequent releases of the software.

  The database interface of the system (interfacing to the Oracle server) uses the *cursor* model of SQL operation. A cursor is first declared, then opened, one or more records retrieved from the database and finally closed. Like any other resource, the number of open cursors that an application may have is limited based on Oracle configuration parameters. The system would correctly close cursors when valid results were obtained from the database or on error, but would not close the cursor when there were no results. This resulted in failure in execution of SQL statements. This problem was fixed in the RDBMS Interface library.

3.1.2 CIP/IMS Facility

The CIP/IMS Facility is a provider gateway that converts incoming CIP requests into the IMS protocol, also mapping attributes from CIP into IMS. The generated ODL request is then forwarded to one or more IMS Servers for execution. Returned results are collated, converted back into CIP format and returned to the user. The exported and imported interfaces of the CIP/IMS facility are shown in Table 3.1.2-1 and Table 3.1.2-2 respectively [IN-2].

The core components of the CIP/IMS Facility [IN-2] are:

- The Retrieval Manager
- The IMS Search Handler and
- The Valids Manager

The Retrieval Manager (RM) is the front-end process of the CIP/IMS Facility. It communicates with the CIP/IMS Translator by sending requests and request responses. The CIP/IMS Facility manages these requests and translates them into IMS format. It writes these requests and signals the IMS Search Handler via socket communication that the request file has been written. When

the Retrieval Manager asks for results, the translator looks for stored ODL results as written by the Search Handler, and translates these IMS results into CIP structures.

The Search Handler is responsible for routing the IMS request originating at the Translator to IMS catalogs. Requests can be sent to multiple IMS servers and the Search handler then waits for the results from these multiple requests to return.

The Valids Manager periodically downloads IMS valids definitions and addresses for existing IMS catalogs from an IMS FTP server. It then interacts with the Operator and the MCF Data Dictionary to see if there are new valids to be processed. If there are new valids, it generates collection data for these new valids and submits them to the Advert & Announcement server, which makes these new collections visible to the user community.

Table 3.1.2-1. CIP/IMS Facility Exported Interfaces

Exported Interface	Purpose	Used By
Notify Update	Notifies the CIP/IMS Facility that a new data dictionary input file has been delivered	MCF Data Dictionary
Transfer Valids	Transfers a new datafile to the MCF/Data Dictionary	MCF Data Dictionary

Table 3.1.2-2. CIP/IMS Facility Imported Interfaces

Server	Imported Interface Description
MCF Data Dictionary	Suggest updates to the data dictionary. The interface also allows the MCF Data Dictionary Server to provide new information to the CIP/IMS Facility.
Advert & Announcement	Submit new collections, modified collections and deletions.

Note that the purpose of the CIP/IMS Facility is to interface with one or more IMS servers, but in the context of the ICS/ECS Gateway is used to interface only with ECS servers i.e. the V0 Gateway.

3.1.2.1 CIP/IMS Facility Configuration & Installation

Configuration of the CIP/IMS Facility included installation of an entry for the Data Center (V0 Gateway instance) that was to be connected to, in a configuration file. The entry required the specification of the V0 Gateway's Data Center ID, Data Center Name, Machine Name and Port number. In addition, entries for the ECS Collection Level Valids was also required to be entered in a configuration file.

3.1.3 V0 Gateway

The V0 Gateway provides interoperability between ECS and the V0 IMS system for inventory searches, browse requests and product orders. It is the primary interface point into the ECS system (as of the current writing). Queries are passed between the V0 IMS system and the V0

Gateway processes using the Object Description Language (ODL). In the context of the ICS/ECS Gateway, the primary client of the V0 Gateway is the CIP/IMS Facility.

The V0 Gateway translates ODL requests into the format expected by ECS Science Data Servers. Since the IMS system uses different attributes to describe data collections within it's data archive, the V0 Gateway must translate those attributes into attributes defined in the ECS system. In order to perform the translation, the V0 Gateway uses the data collection attribute and valid keyword mapping information contained within the ECS Data Dictionary in order to translate the V0 attributes into their equivalent attributes in ECS. Incoming requests are mapped and sent to the ECS Science Data Server for execution.

3.1.3.1 V0 Gateway Configuration & Installation

The V0 Gateway was pre-configured as part of the ECS set of development servers. All that was required was to obtain a machine name and a port number on which the Gateway was listening for incoming requests. The V0 Gateway was configured to access ECS Science Data Servers and other subsystems (MSS, Data Dictionary), in order to process incoming ODL requests.

3.2 IMS/ICS Gateway

The following components constitute the IMS/ICS Gateway.

- EOS Data Gateway
- IMS/ICS Translator
- INFEO

These subsystems are described in the following sections

3.2.1 EOS Data Gateway

Version 2.3.2 of the EOSDIS Data Gateway was used as a test facility to access the IMS/ICS Translator. The installation and configuration of the EOS Data Gateway is described in [EG-1].

3.2.2 IMS/ICS Translator

The internal design of the IMS/ECS Translator is described in [IG-1]. The major configuration item required by the Translator was the set of mapping tables that mapped IMS attributes into CIP attributes.

3.2.3 INFEO

Configuration and installation of the INFEO system is described in Section 3.1.1 above.

This page intentionally left blank.

4. Lessons Learned and Future Work

4.1 Lessons Learned

4.1.1 Prototype Process

Development of the prototypes followed a semi-formal process starting with a design review, followed by issue of a design document of the subsystem being built. The design documents followed the OMT notation of Object Oriented System design. The design documents were circulated among a wide array of interested people with their comments incorporated, before formal development took place. This helped, both in our understanding of the complexities of building the system as well as keeping the customer involved during the entire process. During development, an internal coding standard was followed as well as making use of configuration management tools to keep track of changes and enable sharing of code during parallel development, whenever possible.

4.1.2 Mismatch between documented system and as-builts

There was a frequent problem in interfacing with systems that did not match their published documents. While it is true that implementations of large software systems such as ECS frequently diverge from their design, it was especially difficult for us to access ECS systems due to the lack of support and due to frequently changing interfaces.

4.1.3 Performance Issues

Performance issues were never seriously considered during the development of the prototypes, but serious performance issues do exist. There are many areas that need to be looked at including reduction of the number of data paths/connections to servers and tuning of databases. In an operational setting, performance improvement will have to be made.

4.1.4 System Automation

Many areas of configuration need human intervention. These include analysis of providers data, attribute mapping, database schema mapping etc. In an operational setting, as much of these tasks will need to be automated, either as a set of scripts that can be run or through database forms that will make administration of the system easier.

4.1.5 Firewall Issues

The entire prototype system was developed inside the company firewall, for security purposes. While this was satisfactory in a prototypal setting, in an operational setting, many security issues need to be considered before the system and all it’s servers can be installed outside the firewall.

4.1.6 Operational Transition

The transition from a prototyping stage to an operational stage proved to be an extremely hard task. The primary reason being that in order to be operational, some components of the system needed to be integrated into the ECS system. This proved to be an impossible task, given the time and schedule constraints of the ECS project. This also led to the re-design of some of the gateways, in order to insulate them from having to deal with ECS integration issues, such as moving to the IMS Facility approach in the ICS/ECS Gateway, rather than implementing the more straight forward (and efficient) Catalog Translator approach, in talking to the ECS Science Data Server.

4.2 Future Work

The following list of items describes possible follow on work resulting from the prototyping effort done so far.

- Implementation of a Catalog Translator as designed in [DD-1]. This method of ECS access will provide distinct performance advantages as the number of protocol conversions is reduced, as is the number of servers traversed. In addition, a wider range of attributes are available for catalog searching, since the attribute map now would only have to consider CIP and ECS attributes. Without the intermediate IMS attribute set, a much richer set of attributes are available to the user.

- Implementation of an Order Handling System (OHS). The ICS/ECS Gateway is incomplete without facilities to order data sets. The design for an OHS Translator is described in [OH-1].

- Upgrade of INFEO to CIP Version 2.4. This effort is currently underway by the developers of the INFEO system and the installed system will need to be upgraded when the revised software becomes available.

- In order to make the system operational, the Monitoring and Control Facility (MCF) of INFEO needs to configured and installed.

- Provision of an automated facility for extracting collection level metadata from the ECS system and populating the INFEO Collections database. A set of tools/scripts/database forms will be needed to automate this process.

- The delivery of the IMS/ICS Translator needs to be completed with the addition of *directory search* capability. The current implementation only addresses *inventory search* capability.

Abbreviations and Acronyms

API	Application Programming Interface
CIP	Catalog Interoperability Protocol
COTS	Commercial Off the Shelf
ECS	EOSDIS Core System
ICS	Interoperable Catalog System
IMS	Information Management System
INFEO	Information about Earth Observation
MWND	Middleware Node
ODL	Object Description Language
RM	Retrieval Manager
WWW	World Wide Web

This page intentionally left blank.